



# Riemannian geometry learning for disease progression modelling

Maxime Louis, Raphaël Couronné, Igor Koval, Benjamin Charlier, Stanley Durrleman

## ► To cite this version:

Maxime Louis, Raphaël Couronné, Igor Koval, Benjamin Charlier, Stanley Durrleman. Riemannian geometry learning for disease progression modelling. 2019. hal-02079820v2

**HAL Id: hal-02079820**

**<https://hal.science/hal-02079820v2>**

Preprint submitted on 17 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Riemannian geometry learning for disease progression modelling

Maxime Louis<sup>1,2</sup>, Raphaël Couvonné<sup>1,2</sup>, Igor Koval<sup>1,2</sup>, Benjamin Charlier<sup>1,3</sup>,  
and Stanley Durrleman<sup>1,2</sup>

<sup>1</sup> Sorbonne Universités, UPMC Univ Paris 06, Inserm, CNRS, Institut du cerveau et de la moelle (ICM)

<sup>2</sup> Inria Paris, Aramis project-team, 75013, Paris, France

<sup>3</sup> Institut Montpellierain Alexander Grothendieck, CNRS, Univ. Montpellier

**Abstract.** The analysis of longitudinal trajectories is a longstanding problem in medical imaging which is often tackled in the context of Riemannian geometry: the set of observations is assumed to lie on an a priori known Riemannian manifold. When dealing with high-dimensional or complex data, it is in general not possible to design a Riemannian geometry of relevance. In this paper, we perform Riemannian manifold learning in association with the statistical task of longitudinal trajectory analysis. After inference, we obtain both a submanifold of observations and a Riemannian metric so that the observed progressions are geodesics. This is achieved using a deep generative network, which maps trajectories in a low-dimensional Euclidean space to the observation space.

**Keywords:** Riemannian geometry · Longitudinal progression · Medical Imaging

## 1 Introduction

The analysis of the longitudinal aspects of a disease is key to understand its progression as well as to design prognosis and early diagnostic tools. Indeed, the time dynamic of a disease is more informative than static observations of the symptoms, especially for neuro-degenerative diseases whose progression span over years with early subtle changes. More specifically, we tackle in this paper the issue of disease modelling: we aim at building a time continuous reference of disease progression and at providing a low-dimensional representation of each subject encoding his position with respect to this reference. This task must be achieved using longitudinal datasets, which contain repeated observations of clinical measurements or medical images of subjects over time. In particular, we aim at being able to achieve this longitudinal modelling even when dealing with very high-dimensional data.

The framework of Riemannian geometry is well suited for the analysis of longitudinal trajectories. It allows for principled approaches which respect the nature of the data -e.g. explicit constraints- and can embody some a priori knowledge. When a Riemannian manifold of interest is identified for a given

type of data, it is possible to formulate generative models of progression on this manifold directly. In [3, 10, 14], the authors propose a mixed-effect model which assumes that each subject follows a trajectory which is parallel to a reference geodesic on a Riemannian manifold. In [16], a similar approach is constructed with a hierarchical model of geodesic progressions. All these approaches make use of a predefined Riemannian geometry on the space of observations.

A main limitation of these approaches is therefore the need of this known Riemannian manifold to work on. It may be possible to coin a relevant Riemannian manifold in low-dimensional cases and with expert knowledge, but it is nearly impossible in the more general case of high-dimensional data or when multiple modalities are present. Designing a Riemannian metric is in particular very challenging, as the space of Riemannian metrics on a manifold is vastly large and complex. A first possibility, popular in the literature, is to equip a submanifold of the observation space with the metric induced from a metric on the whole space of observations –e.g.  $\ell^2$  on images. However, we argue here that this choice of larger metric is arbitrary and has no reason to be of particular relevance for the analysis at hand. Another possibility is to consider the space of observations as a product of simple manifolds, each equipped with a Riemannian metric. This is only possible in particular cases, and even so, the product structure constrains the shapes of the geodesics which need to be geodesics on each coordinate. Other constructions of Riemannian metrics exist in special cases, but there is no simple general procedure. Hence, there is a need for data-driven metric estimation.

A few Riemannian metric learning approaches do exist in the literature. In [6], the authors propose to learn a Riemannian metric which is defined by interpolation of a finite set of tensors, and they optimize the tensors so as to separate a set of data according to known labels. This procedure is intractable as the dimension of the data increases. In [1] and in [17], the authors estimate a Riemannian metric so that an observed set of data maximizes the likelihood of a generative model. Their approaches use simple forms for the metric. Finally, in [12], the authors show how to use transformation of the observation space to pull-back a metric from a given space back to the observation space, and give a density criterion for the obtained metric and the data.

We propose in this paper a new approach to learn a smooth manifold and a Riemannian metric which are adapted to the modelling of disease progression. We describe each subject as following a straight line trajectory parallel to a reference trajectory in a low-dimensional latent space  $\mathcal{Z}$ , which is mapped onto a submanifold of the observation space using a deep neural network  $\Psi$ , as seen in [15]. Using the mapping  $\Psi$ , the straight line trajectories are mapped onto geodesics of the manifold  $\Psi(\mathcal{Z})$  equipped with the push-forward of the Euclidean metric on  $\mathcal{Z}$ . After inference, the neural network parametrizes a manifold which is close to the set of observations and a Riemannian metric on this manifold which is such that subjects follow geodesics on the obtained Riemannian manifold, which are all parallel to a common reference geodesic in the sense of [14]. This construction is motivated by the theorem proven in Appendix giving mild conditions under which there exists a Riemannian metric such that a family of

curves are geodesics. Additionally, this particular construction of a Riemannian geometry allows very fast computations of Riemannian operations, since all of them can be done in closed form in  $\mathcal{Z}$ .

Section 2 describes the Riemannian structure considered, the model as well as the inference procedure. Section 3 shows the results on various features extracted from the ADNI data base [7] and illustrates the advantages of the method compared to the use of predefined simple Riemannian geometries.

## 2 Propagation model and deep generative models

### 2.1 Push-forward of a Riemannian metric

We explain here how to parametrize a family of Riemannian manifolds. We use deep neural networks, which we view as non-linear mappings, since they have the advantage of being flexible and computationally efficient.

Let  $\Psi_w : \mathbf{R}^d \mapsto \mathbf{R}^D$  be a neural network function with weights  $w$ , where  $d, D \in \mathbf{N}$  with  $d < D$ . It is shown in [15] that if the transfer functions of the neural network are smooth monotonic and the weight matrices of each layer are of full rank, then  $\Psi$  is differentiable and its differential is of full rank  $d$  everywhere. Consequently,  $\Psi(\mathbf{R}^d) = \mathcal{M}_w$  is locally a  $d$ -dimensional smooth submanifold of the space  $\mathbf{R}^D$ . It is only locally a submanifold:  $\mathcal{M}_w$  could have self intersections since  $\Psi_w$  is in general not one-to-one. Note that using architectures as in [8] would ensure by construction the injectivity of  $\Psi_w$ .

A Riemannian metric on a smooth manifold is a smoothly varying inner product on the manifold tangent space. Let  $g$  be a metric on  $\mathbf{R}^d$ . The push-forward of  $g$  on  $\mathcal{M}_w$  is defined by, for any smooth vector fields  $X, Y$  on  $\Psi_w(\mathbf{R}^d)$ :

$$\Psi_w^*(g)(X, Y) := g((\Psi_w)_*(X), (\Psi_w)_*(Y))$$

where  $(\Psi_w)_*(X)$  and  $(\Psi_w)_*(Y)$  are the pull-back of  $X$  and  $Y$  on  $\mathbf{R}^d$  defined by  $(\Psi_w)_*(X)(f) = X(f \circ \Psi_w^{-1})$  for any smooth function  $f : \mathbf{R}^d \rightarrow \mathbf{R}$ , and where  $\Psi_w^{-1}$  is a local inverse of  $\Psi_w$ , which exists by the local inversion theorem.

By definition,  $\Psi_w$  is an isometry mapping a geodesic in  $(\mathbf{R}^d, g)$  onto a geodesic in  $(\mathcal{M}_w, \Psi_w^*(g))$ . Note that the function  $\Psi_w$  parametrizes both a submanifold  $\mathcal{M}_w$  of the space of observations and a metric  $\Psi_w^*(g)$  on this submanifold. In particular, there may be weights  $w_1, w_2$  for which the manifolds  $\mathcal{M}_{w_1}, \mathcal{M}_{w_2}$  are the same, but the metrics  $\Psi_{w_1}^*(g), \Psi_{w_2}^*(g)$  are different.

In what follows, we denote  $g_w = \Psi_w^*(g)$  the push-forward of the Euclidean metric  $g$ . Since  $(\mathbf{R}^d, g)$  is flat, so is  $(\mathcal{M}_w, g_w)$ . This neither means that  $\mathcal{M}_w$  is flat for the induced metric from the Euclidean metric on  $\mathbf{R}^D$  nor that the obtained manifold is Euclidean (ruled surfaces like hyperbolic paraboloid are flat still non euclidean). A study of the variety of Riemannian manifolds obtained under this form would allow to better understand how vast or limiting this construction is.

## 2.2 Model for longitudinal progression

We denote here  $(y_{ij}, t_{ij})_{j=1, \dots, n_i}$  the observations and ages of the subject  $i$ , for  $i \in \{1, \dots, N\}$  where  $N \in \mathbf{N}$  is the number of subjects and  $n_i \in \mathbf{N}$  is the number of observation of the  $i$ -th subject. The observations lie in a  $D$ -dimensional space  $\mathcal{Y}$ . We model each individual as following a straight trajectory in  $\mathcal{Z} = \mathbf{R}^d$  with  $d \in \mathbf{N}$ :

$$l_i(t) = e^{\eta_i(t - \tau_i)} \mathbf{e}_1 + \sum_{j=2}^d b_i^j \mathbf{e}_j \quad (1)$$

where  $(\mathbf{e}_1, \dots, \mathbf{e}_d)$  is the canonical basis of  $\mathbf{R}^d$ .

With this writing, on average, the subjects follow a trajectory in the latent space given by the direction  $\mathbf{e}_1$ . To account for inter-subject differences in patterns of progression, each subject follows a parallel to this direction in the direction  $\sum_{j=2}^d b_i^j \mathbf{e}_j$ . Finally, we reparametrize the velocity of the subjects in the  $\mathbf{e}_1$  direction using  $\eta_i$  which encodes for the pace of progression and  $\tau_i$  which is a time shift. This writing is so that  $l_i(\tau_i)$  is in  $\text{Vec}(\mathbf{e}_2, \dots, \mathbf{e}_d)$ , the set of all possible states at the time  $\tau_i$ . Hence, after inference, all the subjects progression should be aligned with similar values at  $t = \tau_i$ . We denote, for each subject  $i$ ,  $\varphi_i = (\eta_i, \tau_i, b_i^2, \dots, b_i^d) \in \mathbf{R}^{d+1}$ .  $\varphi_i$  is a low-dimensional vector which encodes the progression of the subject.

As shown above, we map  $\mathcal{Z}$  to  $\mathcal{Y}$  using a deep neural network  $\Psi_w$ . The subject reconstructed trajectories  $t \mapsto y_i(t) = \Psi_w(l_i(t))$  are geodesics in the submanifold  $(\mathcal{M}_w, g_w)$ . The geodesics are parallel in the sense of [14] and [16]. Note that the apparently constrained form of latent space trajectories (1) is not restrictive: the family of functions parametrized by the neural network  $\Psi_w$  allows to curve and move the image of the latent trajectories in the observation space, and for example to align the direction  $\mathbf{e}_1$  with any direction in  $\mathcal{Y}$ .

## 2.3 Encoding the latent variables

To predict the variables  $\varphi$  for a given subject, we use a recurrent neural-network (RNN), which is to be thought as an encoder network. As noted in [4, 5], the use of a recurrent network allows to work with sequences which have variable lengths. This is a significant advantage given the heterogeneity of the number of visits in medical longitudinal studies. In practice, the observations of the subject are not regularly spaced in time. To allow the network to adapt to this, we provide the ages of the visit at each update of the RNN.

We use an Elman network, which has a hidden state  $h \in \mathbf{R}^H$  with  $H \in \mathbf{N}$ , initialized to  $h_0 = 0$  and updated along the sequence according to  $h_k = \rho_h(W_h y_{ik} + U_h h_{k-1} + b_h)$  and the final value predicted by the network after a sequence of length  $f \in \mathbf{N}$  is  $\varphi = W_\varphi \rho_\varphi(h_f) + b_\varphi$  where  $\rho_\varphi$  and  $\rho_h$  are activation functions and  $W_h, U_h, W_\varphi, b_h, b_\varphi$  are the weights and biases of the network. We denote  $\theta = (W_h, U_h, W_\varphi, b_\varphi)$  the parameters of the encoder.

When working with scalar data, we use this architecture directly. When working with images, we first use a convolutional neural network to extract relevant

features from the images which are then fed to the RNN. In this case, both the convolutional network and the recurrent network are trained jointly by back-propagation. Figure 1 summarizes the whole procedure.

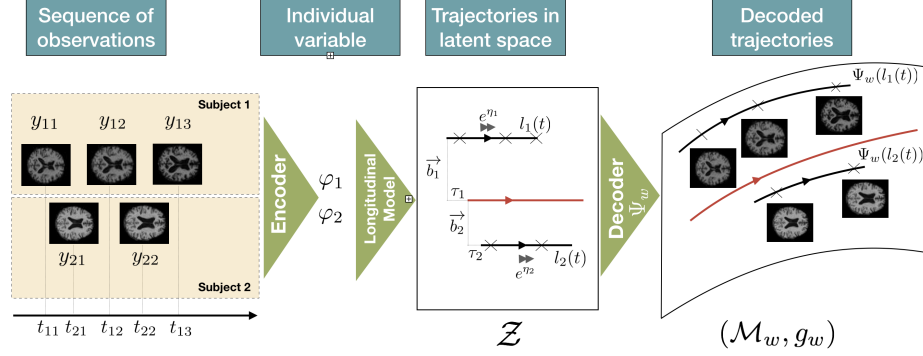


Fig. 1: The observed sequences are encoded into latent trajectories, which are then decoded into geodesics on a submanifold of the observation space.

## 2.4 Regularization

To impose some structure in the latent space  $\mathcal{Z}$ , we impose a regularization on the individual variables  $\varphi_i = (\eta_i, \tau_i, b_i^2, \dots, b_i^d)$ . The regularization cost used is:

$$r(\eta, \tau, b^2, \dots, b^d) = \frac{(\eta - \bar{\eta})^2}{\sigma_\eta^2} + \frac{(\tau - \bar{\tau})^2}{\sigma_\tau^2} + \sum_{j=2}^d (b^j)^2 \quad (2)$$

This regularization requires the individual variables  $\eta$  and  $\tau$  to be close to mean values  $\bar{\tau}, \bar{\eta}$ . The parameters  $\bar{\eta}, \bar{\tau}$  are estimated during the inference.  $\sigma_\eta > 0$  is fixed but the estimation of  $\bar{\eta}$  allows to adjust the typical variation of  $\eta$  with respect to the mean pace  $\bar{\eta}$ , while the neural network  $\Psi_w$  adjusts accordingly the actual velocity in the observation space in the  $\mathbf{e}_1$  direction.  $\sigma_\tau$  is set to the empirical standard deviation of the time distribution  $(t_{ij})_{ij}$ , meaning that we expect the delays between subjects to be of the same order as the standard deviation of the visit ages.

## 2.5 Cost function and inference

Overall, we optimize the cost function:

$$c(\theta, w, \bar{\eta}, \bar{\tau}) = \sum_i \sum_j \frac{\|y_i(t_{ij}) - y_{ij}\|_2^2}{\sigma^2} + \sum_i r(\varphi_i) \quad (3)$$

where  $\sigma > 0$  is a parameter controlling the trade-off reconstruction/regularity.

The first term contains the requirements that the geometry  $(\mathcal{M}_w, g_w)$  be adapted to the observed progressions since it requires geodesics  $y_i(t)$  to be good reconstructions of the individual trajectories. As shown in the Appendix, there exists solutions to problems of this kind: under mild conditions there exists a metric on a Riemannian manifold that the subjects progressions are geodesics. But this is only a partial constraint: there is a whole class of metrics which have geodesics in common (see [13] for the analysis of metrics which have a given family of trajectories as geodesics).

We infer the parameters of the model by stochastic gradient descent using the Adam optimizer [9]. After each batch of subjects  $B$ , we balance regularity and reconstruction by imposing a closed-form update on  $\sigma$ :

$$\sigma^2 = \frac{1}{N_B D} \sum_{i \in B} \sum_{j=1}^{N_i} \|\Psi_w(l_i(t_{ij})) - y_{ij}\|_2^2 \quad (4)$$

where  $N_B = \sum_{i \in B} N_i$  is the total number of observations in the batch  $b$ . This automatic update of the trade-off criterion  $\sigma$  is inspired from Bayesian generative models which estimate the variance of the residual noise, as in e.g. [14, 20].

### 3 Experimental results

The neural network architectures and the source code for these experiments is available at [gitlab.icm-institute.org/maxime.louis/unsupervised\\_geometric\\_longitudinal](https://gitlab.icm-institute.org/maxime.louis/unsupervised_geometric_longitudinal), tag IPMI 2019. For all experiments, the ages of the subjects are first normalized to have zero mean and unit variance. This allows the positions in the latent space to remain close to 0. We set  $\sigma_\eta = 0.5$  and initialize  $\bar{\eta}$  to 0.

#### 3.1 On a synthetic set of images

To validate the proposed methodology, we first perform a set of experiments on a synthetic data set. We generate  $64 \times 64$  gray level images of a white cross on a black background. Each cross is described by the arms length and angles. We prescribe a mean scenario of progression for the arm lengths and sample the arm angles for each subject from a zero-centered normal distribution. Figure 2 shows subjects generated along this procedure. Note that with this setting, an image varies smoothly with respect to the arms lengths and angles and hence the whole set of generated images is a smooth manifold.

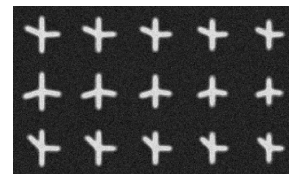


Fig. 2: Each row represents a synthetic subject.

We generate 10 training sets of 150 subjects and 10 test sets of 50 subjects. The number of observation of each subject is randomly sampled from a Poisson distribution with mean 5. The times at which the subject

are observed are equally spaced within a randomly selected time window. We add different level of white noise on the images. We then run the inference on the 10 training sets for each level of noise. We set the dimension of the latent space  $\mathcal{Z}$  to 3 for all the experiments.

For each run, we estimate the reconstruction error on both training set and test set, as well as the reconstruction error to the de-noised images, which were not seen during training. Results are shown on Figure 3.

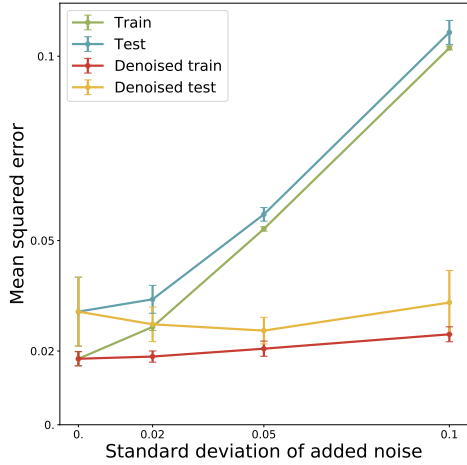


Fig. 3: Reconstruction error on train and test sets and on denoised train and test sets, unseen during training.

The model generalizes well to unseen data and successfully denoises the images, with a reconstruction error on the denoised images which does not vary with the scale of the added white noise. This means that the generated manifold of images is close to the manifold on which the original images lie. Besides, as shown on Figure 4, the scenario of progression along the  $e_1$  direction is well captured, while orthogonal directions  $e_2, e_3$  allow to change the arm positions. Finally, we compare the individual variables ( $b_i^2, b_i^3$ ) to the known arms angles which were used to generate the images. Figure 5 shows the results: the latent space is structured in a way that is faithful to the original arm angles.

### 3.2 On cognitive scores

We use the cognitive scores grading the subjects memory, praxis, language and concentration from the ADNI database as in [14]. Each score is renormalized to vary between 0 and 1, with large values indicating poor performances for the task. Overall, the data set consists of 223 subjects observed 6 times on average over 2.9 years. We perform a 10-fold estimation of the model. The measured mean squared reconstruction error is  $0.079 \pm 1.1e-3$  on the train sets, while it is of  $0.085 \pm 1.5e-3$  for the test sets. Both are close to the uncertainty in the estimation of these cognitive scores [18]. This illustrates the ability of the model to generalize to unseen observations. First, this indicates that  $M_w$  is a submanifold which is close to all the observations. Second, it indicates how relevant the learned Riemannian metric is, since unobserved subject trajectories are very close to geodesics on the Riemannian manifold.

Figure 6 shows obtained average trajectories  $t \mapsto \Psi_w(e^{\bar{\eta}}e_0(t - \bar{\tau}))$  for a 10-fold estimation of the model on the data set, with  $\mathcal{Z}$  dimension set to 2. All of these trajectories are brought back to a common time reference frame using the estimated  $\bar{\tau}$  and  $\bar{\eta}$ . All average trajectories are very similar, underlining the stability of the proposed method. Note that despite the small average observation



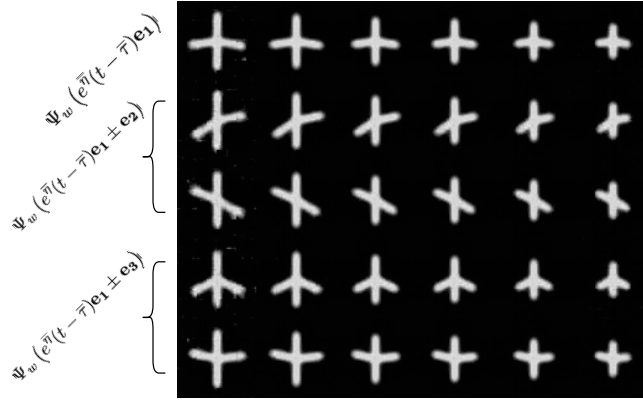


Fig. 4: First row is  $t \mapsto \Psi_\theta(e_1 t)$ . Following rows are  $t \mapsto \Psi_\theta(e_1 t + e_i)$  for  $i \in \{2, 3\}$ . These parallel directions of progression show the same arm length reduction with different arm positions.

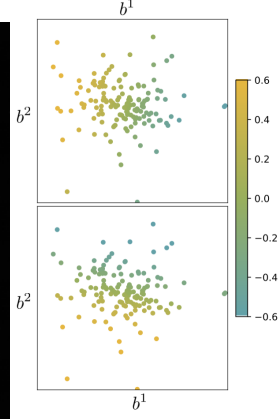


Fig. 5: Individual variables  $b_i^1$  and  $b_i^2$  colored by left (top) and right (bottom) arm angle.

time of the subjects, the method proposed here allows to robustly obtain a mean scenario of progression over 30 years. Hence, despite all the flexibility that is provided through the different neural networks and the individual parameters, the model still exhibits a low variance.

We compare the results to the mean trajectory estimated by the model [14], which is shown on Figure 8. Both cases recover the expected order of onset of the different cognitive symptoms in the disease progression. Note that with

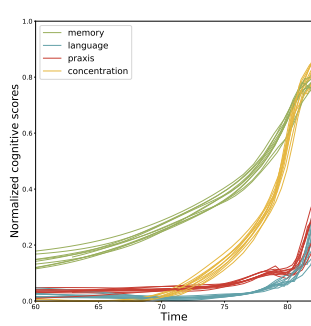


Fig. 6: Learned main progression of the cognitive scores, for the 10-fold estimation.

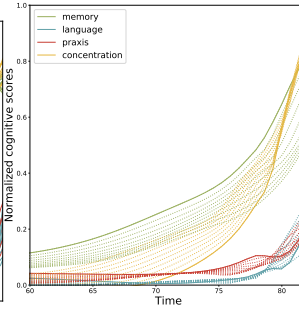


Fig. 7: Mean geodesic of progression and parallel variations  $t \mapsto \Psi(e\bar{\eta}e_0(t - \bar{\tau}) + \lambda e_1)$  for varying  $\lambda$ .

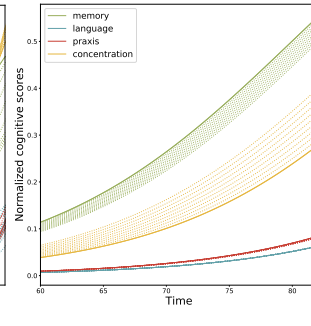


Fig. 8: Mean geodesic of progression and parallel variations for the model with user-defined metric.

our model the progression of the concentration score is much faster than the progression of the memory score, although it starts later: this type of behaviour is not allowed with the model described in [14] where the family of possible trajectories is much narrower. Indeed, because it is difficult to craft a relevant Riemannian metric for the data at hand, the authors modelled the data as lying on a product Riemannian manifold. In this case, a geodesic on the product manifold is a product of geodesics of each manifold. This strongly restricts the type of dynamics spanned by the model and hence gives it a high bias.

The use of the product manifold also has an impact on the parallel variations around the mean scenario: they can only delay and slow/accelerate one of the component with respect to another, as shown on Figure 8. Figure 7 illustrates the parallel variations  $\Psi(\bar{\alpha}e_0(t - \bar{\tau}) + \mathbf{e}_1)$  one can observe with the proposed model. The variation is less trivial since complex interactions between each features are possible. In particular, the concentration score varies more in the early stages of the disease than in the late stages.

**The individual variables  $\varphi_i$ .** To show that the individual variables  $\varphi_i$  did capture information regarding the disease progression, we compare the distribution of the  $\tau_i$  between subjects who have at least one APOE4 allele of the APOE gene -an allele known to be implicated in Alzheimer’s disease- and subjects which have no APOE4 allele of this gene. We perform a Mann-Whitney test on the distributions to see if they differ. For all folds, a p-value lower than 5% is obtained. For all folds, carriers have a larger  $\tau$  meaning that they have an earlier disease onset than non-carriers. This is in accordance with [2]. Similarly, women have on average an earlier disease onset for all folds, in accordance with [11].

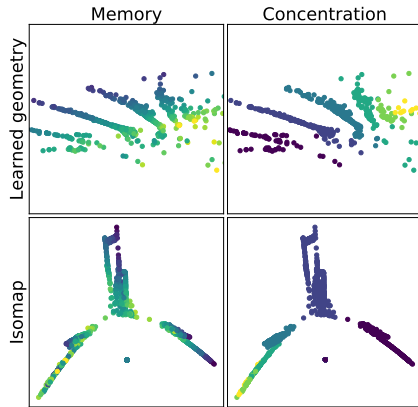


Fig. 9: Top (resp. Bottom) latent positions (resp. Isomap embedding) of the observations colored by memory and concentration score.

**A closer look at the geometry** We look at the obtained Riemannian geometry by computing the latent position best mapped onto each of the observation by  $\Psi_w$ . We then plot the obtained latent positions to look at the structure of the learned Riemannian manifold. We compare the obtained structure with a visualisation of the structure induced by the  $\ell^2$  on the space of observations produced by Isomap [19]. Isomap is a manifold learning technique which attempts to reconstruct in low dimensions the geodesic distances computed from a set of observations. The results are shown on Figure 9. The geometry obtained after inference is clearly much more suited for disease progression modelling. Indeed, the  $\mathbf{e}_1$  direction does correspond to typical increases in the dif-

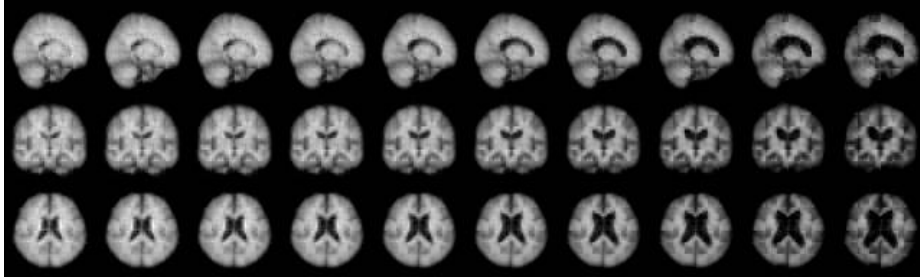


Fig. 10:  $t \mapsto \Psi_\theta(e_0 t)$  on the MRI dataset. The growth of the ventricles, characteristic of aging and Alzheimer’s disease is clearly visible.

ferent scores. The induced metric is not as adapted. This highlights the relevance of the learned geometry for disease modelling.

### 3.3 On anatomical MRIs

We propose here an estimation of the model on 3D MRIs preprocessed from the ADNI database, to check the behaviour of the proposed method in high dimension. We select subjects which ultimately convert to Alzheimer’s disease. We obtain 325 subjects and a total of 1996 MRIs which we affinely align on the Colin-27 template and resample to  $64 \times 64 \times 64$ . We run a 5-fold estimation of the model with  $\dim \mathcal{Z} = 10$ , using a GPU backend. We obtain a train mean squared error of  $0.002 \pm 1e - 5$  and a test mean squared error of  $0.0024 \pm 3e - 5$ . Figure 10 shows one of the learned mean trajectory.

Once the model is estimated, we compare the distributions of the pace of progressions  $\eta_i$  between the individuals who have at least one APOE4 allele of the APOE gene and the individuals who have no APOE4 allele. For all 5 folds, the distributions of the paces of progression significantly differ, with p-values lower than 5% and in each case, the APOE4 carriers have a greater pace of progression, in accordance with [2]. The same analysis between the individuals who have two APOE4 allele versus the individuals which have at maximum one APOE4 allele shows a significative difference for all folds for the  $\tau$  variable: the APOE4 carriers have an earlier disease onset, as shown in [11]. This analysis further shows the value of the individual variables  $\varphi_i$  learned for each subject.

## 4 Conclusion

We presented a way to perform Riemannian geometry learning in the context of longitudinal trajectory analysis. We showed that we could build a local Riemannian submanifold of the space of observations which is so that each subject follows a geodesic parallel to a main geodesic of progression on this manifold. We illustrated how the encoding of each subject into these trajectories is informative of the disease progression. The latent space  $\mathcal{Z}$  built in this process is a

low-dimensional description of the disease progression. There are several possible continuations of this work. First, there is the possibility to conduct the same analysis on multiple modalities of data simultaneously. Then, after estimation, the latent space  $\mathcal{Z}$  could be useful to perform classification and prediction tasks.

*Acknowledgements* This work has been partially funded by the European Research Council (ERC) under grant agreement No 678304, European Union’s Horizon 4582020 research and innovation programme under grant agreement No 666992, and the program ”Investissements d’avenir” ANR-10-IAIHU-06.

## References

1. Arvanitidis, G., Hansen, L.K., Hauberg, S.: A locally adaptive normal distribution. In: Advances in Neural Information Processing Systems. pp. 4251–4259 (2016)
2. Bigio, E., Hynan, L., Sontag, E., Satumtira, S., White, C.: Synapse loss is greater in presenile than senile onset alzheimer disease: implications for the cognitive reserve hypothesis. *Neuropathology and applied neurobiology* **28**(3), 218–227 (2002)
3. Bône, A., Colliot, O., Durrleman, S.: Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. arXiv preprint arXiv:1803.10119 (2018)
4. Cui, R., Liu, M., Li, G.: Longitudinal analysis for alzheimer’s disease diagnosis using rnn. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. pp. 1398–1401. IEEE (2018)
5. Gao, L., Pan, H., Liu, F., Xie, X., Zhang, Z., Han, J., Initiative, A.D.N., et al.: Brain disease diagnosis using deep learning features from longitudinal mr images. In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. pp. 327–339. Springer (2018)
6. Hauberg, S., Freifeld, O., Black, M.J.: A geometric take on metric learning. In: Advances in Neural Information Processing Systems. pp. 2024–2032 (2012)
7. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **27**(4), 685–691 (2008)
8. Jacobsen, J.H., Smeulders, A., Oyallon, E.: i-revnet: Deep invertible networks. arXiv preprint arXiv:1802.07088 (2018)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Koval, I., Schiratti, J.B., Routier, A., Bacci, M., Colliot, O., Allasonnière, S., Durrleman, S., Initiative, A.D.N., et al.: Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 451–459. Springer (2017)
11. Lam, B., Masellis, M., Freedman, M., Stuss, D.T., Black, S.E.: Clinical, imaging, and pathological heterogeneity of the alzheimer’s disease syndrome. *Alzheimer’s research & therapy* **5**(1), 1 (2013)
12. Lebanon, G.: Learning riemannian metrics. In: Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence. pp. 362–369. Morgan Kaufmann Publishers Inc. (2002)

13. Matveev, V.S.: Geodesically equivalent metrics in general relativity. *Journal of Geometry and Physics* **62**(3), 675–691 (2012)
14. Schiratti, J.B., Allasonniere, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: *Advances in Neural Information Processing Systems*. pp. 2404–2412 (2015)
15. Shao, H., Kumar, A., Fletcher, P.T.: The riemannian geometry of deep generative models. *arXiv preprint arXiv:1711.08014* (2017)
16. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: Hierarchical geodesic models in diffeomorphisms. *International Journal of Computer Vision* **117**(1), 70–92 (2016)
17. Sommer, S., Arnaudon, A., Kuhnelt, L., Joshi, S.: Bridge simulation and metric estimation on landmark manifolds. In: *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*, pp. 79–91. Springer (2017)
18. Standish, T.I., Molloy, D.W., Bédard, M., Layne, E.C., Murray, E.A., Strang, D.: Improved reliability of the standardized alzheimer’s disease assessment scale (sadas) compared with the alzheimer’s disease assessment scale (adas). *Journal of the American Geriatrics Society* **44**(6), 712–716 (1996)
19. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *science* **290**(5500), 2319–2323 (2000)
20. Zhang, M., Fletcher, T.: Probabilistic principal geodesic analysis. In: *Advances in Neural Information Processing Systems*. pp. 1178–1186 (2013)

## A Existence of a Riemannian metric such that a family of curves are geodesics

**Theorem 1.** *Let  $\mathcal{M}$  be a smooth manifold and  $(\gamma_i)_{i \in \{1, \dots, n\}}$  be a family of smooth regular injective curves on  $\mathcal{M}$  which do not intersect. There exists a Riemannian metric  $g$  such that  $\gamma_i$  is a geodesic on  $(\mathcal{M}, g)$  for all  $i \in \{1, \dots, n\}$ .*

*Proof. Open neighborhood of the curves.* Let  $i \in I$ . Since  $\gamma_i$  is injective and regular,  $\gamma_i([0, 1])$  is a submanifold of  $\mathcal{M}$ . By the tubular neighborhood theorem, there exists a vector bundle  $E_i$  on  $\gamma_i([0, 1])$  and a diffeomorphism  $\Phi_i$  from a neighborhood  $U_i$  of the 0-section in  $E_i$  to a neighborhood  $V_i$  of  $\gamma_i([0, 1])$  in  $\mathcal{M}$  such that  $\Phi_i \circ 0_{E_i} = i_i$  where  $i_i$  is the embedding of  $\gamma_i([0, 1])$  in  $\mathcal{M}$ . Without loss of generality, we can suppose  $\gamma_j([0, 1]) \cap U_i = \emptyset$  for all  $j \neq i$ .

*Riemannian metric on the neighborhoods.* To do so, we use the fact that  $E_i$  is diffeomorphic to the normal bundle on the segment  $[0, 1] \in \mathbf{R}$  which is trivial and which we denote  $N$ . Let  $\Psi_i : E_i \mapsto N$  be such a diffeomorphism. Now  $N$  can be equipped with a Riemannian metric  $h_i$  such that  $[0, 1]$  is a geodesic. Using  $\Psi_i$  and  $\Phi_i$ , we can push-forward the metric  $h_i$  to get a metric  $g_i$  on  $U$  which is so that  $\gamma_i$  is a geodesic on  $(U_i, g_i)$ .

*Stitching the metrics with a partition of unity.* For each  $i \in \{1, \dots, n\}$ , pick an open subset  $V_i$  such that  $V_i \subset \bar{V}_i \subset U_i$  and which contains  $\gamma_i([0, 1])$ , and set  $O = \mathcal{M} \setminus (\cup_i \bar{V}_i)$ .  $O$  is open so that  $\mathcal{C} = \{O, U_1, \dots, U_n\}$  is an open cover of  $\mathcal{M}$ .  $O$  can be equipped with a metric  $g_O$  (there always exists a Riemannian metric on a smooth manifold). Finally, we use a partition of the unity  $\rho_O, \rho_1, \dots, \rho_n$  on  $\mathcal{C}$  and set  $g = \rho_O g_O + \sum_i \rho_i g_i$ .  $g$  is a Riemannian metric on  $\mathcal{M}$  as a positive combination of Riemannian metrics. Each  $\gamma_i$  is a geodesic on  $(\mathcal{M}, g)$  by construction.  $\square$

In [13], the authors deal with a more general case but obtain a less explicit existence result. This theorem motivates our approach even if in equation (3) we ask for more: we want the existence a system of coordinates adapted to the progression.